

# Markov Chain Monte Carlo Estimation of Quantiles

James M. Flegal  
 Department of Statistics  
 University of California, Riverside  
 jflegal@ucr.edu

Galin L. Jones <sup>\*</sup>  
 School of Statistics  
 University of Minnesota  
 galin@umn.edu

Ronald C. Neath  
 Department of Mathematics and Statistics  
 Hunter College, City University of New York  
 rneath@hunter.cuny.edu

July 30, 2012

## Abstract

We consider quantile estimation using Markov chain Monte Carlo and establish conditions under which the sampling distribution of the Monte Carlo error is approximately Normal. Further, we investigate techniques to estimate the associated asymptotic variance, which enables construction of an asymptotically valid interval estimator. Finally, we explore the finite sample properties of these methods through examples. An R package, `mcmcse`, makes implementation of the suggested methods easy.

## 1 Introduction

Let  $\pi$  denote a probability distribution having support  $\mathsf{X} \subseteq \mathbb{R}^d$ ,  $d \geq 1$ . If  $W \sim \pi$  and  $g : \mathsf{X} \rightarrow \mathbb{R}$ , set  $V = g(W)$ . We consider estimation of quantiles of the distribution of  $V$ . Specifically, if  $F_V$  denotes the cumulative distribution function of  $V$ , then our goal is to obtain

$$\xi_q := F_V^{-1}(q) = \inf\{v : F_V(v) \geq q\}.$$

In most practically relevant statistical settings it is not possible to calculate  $\xi_q$  directly. For example, a common goal is construction of posterior credible intervals via calculation of quantiles of marginal posterior distributions.

---

<sup>\*</sup>Research supported by the National Institutes of Health.

Our focus is on using Markov chain Monte Carlo (MCMC) methods to approximate  $\xi_q$ . The basic MCMC method entails simulating a Markov chain  $X = \{X_0, X_1, \dots\}$  having invariant distribution  $\pi$ ; the popularity of MCMC largely is due to the ease with which such a simulation can be accomplished (Liu, 2001; Robert and Casella, 2004). Define  $Y = \{Y_0, Y_1, \dots\} = \{g(X_0), g(X_1), \dots\}$ . If we observe a realization of  $X$  of length  $n$  and let  $Y_{n(j)}$  denote the  $j$ th order statistic of  $\{Y_0, \dots, Y_{n-1}\}$ , then a natural estimator of  $\xi_q$  is given by

$$\hat{\xi}_{n,q} := Y_{n(j+1)} \quad \text{where} \quad j \leq nq < j+1. \quad (1)$$

The estimate  $\hat{\xi}_{n,q}$  will be more valuable if we can also assess the unknown Monte Carlo error,  $\hat{\xi}_{n,q} - \xi_q$ . We consider doing this through its approximate sampling distribution.

If  $\{X_0, \dots, X_{n-1}\}$  is a random sample from  $\pi$  and  $F_V$  has a density  $f_V$  positive at  $F_V^{-1}(q)$ , then a classical result is that as  $n \rightarrow \infty$

$$\sqrt{n}(\hat{\xi}_{n,q} - \xi_q) \xrightarrow{d} N(0, q(1-q)/[f_V(\xi_q)]^2).$$

Moreover, it is easy to estimate the variance in the asymptotic distribution and hence construct an asymptotically valid interval estimator for  $\xi_q$ . Our goal is to extend this approach to the setting where  $X$  is a Markov chain and consider its application in the context of MCMC. Unfortunately, the above conditions are no longer sufficient for a CLT; we must replace the assumption of  $X$  being a random sample with a mixing condition on the Markov chain. While we defer the statement of our theorem until Section 2, we note that the required mixing condition is quite weak; specifically, polynomially ergodicity of order 3 will suffice. For now, assume there exists a constant  $\gamma_q^2 > 0$  such that as  $n \rightarrow \infty$

$$\sqrt{n}(\hat{\xi}_{n,q} - \xi_q) \xrightarrow{d} N(0, \gamma_q^2). \quad (2)$$

Note that  $\gamma_q^2$  must account for the serial dependence present in a non-trivial Markov chain and hence is a more complicated quantity than when  $X$  is a random sample.

Suppose we can construct an estimate of  $\gamma_q^2$ , say  $\hat{\gamma}_n^2$ , then an interval estimator of  $\xi_q$  is given by

$$\hat{\xi}_{n,q} \pm t_* \frac{\hat{\gamma}_n}{\sqrt{n}}$$

where  $t_*$  is an appropriate Student's  $t$  quantile. Such intervals, or at least, the *Monte Carlo standard error* (MCSE),  $\hat{\gamma}_n/\sqrt{n}$ , are useful in assessing the reliability of the simulation results as they explicitly describe the level of confidence we have in the reported number of significant figures in  $\hat{\xi}_{n,q}$ . In this sense reporting the MCSE allows us to have as much confidence in the MCMC simulation results as we would if it were possible to simulate a random sample from

$\pi$ . For more on this approach see Flegal et al. (2008), Flegal and Jones (2011), Geyer (2011), Jones et al. (2006) and Jones and Hobert (2001).

We consider two methods for implementing this recipe. The first is based on the method of batch means (BM) while the second is based on regenerative simulation (RS). Batch means is based on the CLT in (2) and works by breaking the chain up into batches that are approximately independent and then operating on these batches. RS is based on simulating an augmented Markov chain  $X'$  which allows breaking the chain into independent and identically distributed parts and hence requires that we establish a slightly different quantile CLT than that in (2). We will show that the BM-based interval is easier to implement, but the RS-based interval requires slightly weaker regularity conditions. Moreover, our simulation results show that both produce effective interval estimates of  $\xi_q$ .

The remainder is organized as follows. We begin in Section 2 by setting our notation and carefully stating our assumptions. We then establish a CLT for the Monte Carlo error and consider how to calculate Monte Carlo standard errors using BM. Next we consider RS and establish an alternative CLT. We go on to show how RS can be used for calculating an MCSE. In the last part of Section 2, we consider alternative methods such as bootstrap methods for time series. In Section 3 we illustrate the use of RS and BM and investigate their finite-sample properties in two examples. Finally, in Section 4 we summarize our results and conclude with a practical recommendation.

## 2 Quantile estimation

Recall that  $\pi$  is a probability distribution having support  $\mathbf{X}$  and let  $\mathcal{B}(\mathbf{X})$  denote the Borel  $\sigma$ -algebra. Throughout we assume  $X$  is a Harris ergodic (that is,  $\pi$ -irreducible, aperiodic and positive Harris recurrent) Markov chain having invariant distribution  $\pi$ .

As we noted in the introduction a mixing condition on  $X$  is one of the sufficient conditions for a CLT. We require some notation before we can describe the mixing condition. For  $n \in \{1, 2, 3, \dots\}$  let the  $n$ -step Markov kernel associated with  $X$  be  $P^n(x, dy)$ . Then if  $A \in \mathcal{B}(\mathbf{X})$  and  $k \in \{0, 1, 2, \dots\}$ ,  $P^n(x, A) = \Pr(X_{k+n} \in A | X_k = x)$ . Let  $\|\cdot\|$  denote the total variation norm. Say  $X$  is *polynomially ergodic of order  $m$*  if there exists a real-valued function  $M : \mathbf{X} \rightarrow \mathbb{R}^+$  with  $E_\pi M < \infty$  and  $m > 0$  such that for all  $X_0 = x$

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)n^{-m}. \quad (3)$$

There is a substantial body of work concerning polynomial ergodicity of MCMC algorithms; see e.g. Douc et al. (2004), Fort and Moulines (2000), Fort and Moulines (2003), Jarner

and Roberts (2002), Jarner and Roberts (2007), Jarner and Tweedie (2003). However, in the MCMC literature more attention has been paid to establishing geometric ergodicity and uniform ergodicity, both of which are stronger than polynomial ergodicity; see, among many others, Hobert (2011), Jones and Hobert (2001), Johnson et al. (2011), Jones et al. (2012), Mengersen and Tweedie (1996), Roberts and Tweedie (1996) and Tierney (1994).

We are now in position to give our first CLT. Recall  $Y = \{Y_0, Y_1, \dots\} = \{g(X_0), g(X_1), \dots\}$ , define  $U(y) = \{U_0(y), U_1(y), \dots\} = \{I(Y_0 \leq y), I(Y_1 \leq y), \dots\}$  and set

$$\sigma^2(y) := \text{Var}_\pi U_0(y) + 2 \sum_{k=1}^{\infty} \text{Cov}_\pi[U_0(y), U_k(y)] . \quad (4)$$

The proof of the following result is easily obtained from Theorem 2 in Jones (2004) and Theorem 2 in Yoshihara (1995).

**Theorem 1.** *Suppose there is  $\epsilon > 0$  such that  $X$  is polynomially ergodic of order  $2.5 + \epsilon$ . If  $F_V$  has a density  $f_V$  positive and bounded in some neighborhood of  $\xi_q$ , then as  $n \rightarrow \infty$  the sum in (4) converges absolutely and if  $\sigma^2(\xi_q) > 0$*

$$\sqrt{n}(\hat{\xi}_{n,q} - \xi_q) \xrightarrow{d} N(0, \sigma^2(\xi_q)/[f_V(\xi_q)]^2) \quad n \rightarrow \infty . \quad (5)$$

To obtain a Monte Carlo standard error we need to estimate the variance of the asymptotic Normal distribution in Theorem 1, that is  $\gamma^2(\xi_q) := \sigma^2(\xi_q)/[f_V(\xi_q)]^2$ . First, we substitute for  $\xi_q^2$  and separately consider estimating  $f_V(\hat{\xi}_{n,q})$  and  $\sigma^2(\hat{\xi}_{n,q})$ .

Consider estimating  $f_V(\hat{\xi}_{n,q})$ —a univariate density estimation problem. Kernel density estimation has been studied extensively in the context of stationary time-series analysis (see e.g. Robinson, 1983) and many existing results are applicable since the Markov chains in MCMC are special cases of strong mixing processes. In our examples we use kernel density estimators with a gaussian kernel, for which there are well known conditions guaranteeing consistent estimation of the density at a point; see e.g. Kim and Lee (2005) and Yu (1993).

The quantity  $\sigma^2(y)$ ,  $y \in \mathbb{R}$  is familiar. Notice that if  $\bar{U}_n(y) := n^{-1} \sum_{i=0}^{n-1} U_i(y)$ , then

$$\sqrt{n}(\bar{U}_n(y) - E_\pi I(Y \leq y)) \xrightarrow{d} N(0, \sigma^2(y)) \quad n \rightarrow \infty$$

by the usual Markov chain CLT for sample means. In this context, estimating  $\sigma^2(y)$  is a well-studied problem and there are an array of methods for doing so; see Geyer (1996), Geyer (2011), Flegal and Jones (2010), Hobert et al. (2002), Jones et al. (2006) and Mykland et al. (1995) among others.

We will use the method of batch means for estimating  $\sigma^2(\hat{\xi}_{n,q})$ . For BM the output is split into batches of equal size. Suppose we obtain  $n = a_n b_n$  iterations  $\{X_0, \dots, X_{n-1}\}$  and

for  $k = 0, \dots, a_n - 1$  define  $\bar{U}_k(\hat{\xi}_{n,q}) := b_n^{-1} \sum_{i=0}^{b_n-1} U_{kb_n+i}(\hat{\xi}_{n,q})$ . The BM estimate of  $\sigma^2(\hat{\xi}_{n,q})$  is

$$\hat{\sigma}_{BM}^2(\hat{\xi}_{n,q}) = \frac{b_n}{a_n - 1} \sum_{k=0}^{a_n-1} \left( \bar{U}_k(\hat{\xi}_{n,q}) - \bar{U}_n(\hat{\xi}_{n,q}) \right)^2. \quad (6)$$

The theoretical and empirical properties of BM methods in MCMC settings have been studied by Flegal et al. (2008), Flegal and Jones (2010), Flegal and Jones (2011) and Jones et al. (2006) who found that  $b_n = \lfloor n^{1/2} \rfloor$  worked well and this is what we will use in our examples.

Letting  $\hat{f}_V(\hat{\xi}_{n,q})$  denote the density estimator of  $f_V(\hat{\xi}_{n,q})$  we estimate of  $\gamma^2(\xi_q)$  with

$$\hat{\gamma}^2(\hat{\xi}_{n,q}) := \frac{\hat{\sigma}_{BM}^2(\hat{\xi}_{n,q})}{\hat{f}_V(\hat{\xi}_{n,q})},$$

yielding an approximate  $100(1 - \alpha)\%$  confidence interval for  $\xi_q$ , with  $z_{\alpha/2}$  an appropriate standard Normal quantile, given by

$$\hat{\xi}_{n,q} \pm z_{\alpha/2} \frac{\hat{\gamma}(\hat{\xi}_{n,q})}{\sqrt{n}}. \quad (7)$$

This approach is implemented in the R package `mcmcse` (Flegal, 2012b) which is used to perform many of the computations in our examples.

## 2.1 Regenerative Simulation

Regenerative simulation (RS) provides an alternative to BM. RS is based on simulating an augmented Markov chain and so Theorem 1 does not apply. We derive an alternative CLT based on RS and consider a natural estimator of the variance in the asymptotic Normal distribution.

Recall  $X$  has  $n$ -step Markov kernel  $P^n(x, dy)$  and suppose the kernel satisfies a one-step *minorization condition*, that is, suppose there exists a function  $s : \mathbf{X} \rightarrow [0, 1]$  with  $E_\pi s > 0$  and a probability measure  $Q$  such that

$$P(x, A) \geq s(x)Q(A) \quad \text{for all } x \in \mathbf{X} \text{ and } A \in \mathcal{B}. \quad (8)$$

We call  $s$  the *small function* and  $Q$  the *small measure*. In this case we can write

$$P(x, dy) = s(x)Q(dy) + (1 - s(x))R(x, dy) \quad (9)$$

where  $R$  is the *residual measure*, given by

$$R(x, dy) = \begin{cases} \frac{P(x, dy) - s(x)Q(dy)}{1 - s(x)} & s(x) < 1 \\ Q(dy) & s(x) = 1. \end{cases} \quad (10)$$

We now have the ingredients for constructing the *split chain*,

$$X' = \{(X_0, \delta_0), (X_1, \delta_1), (X_2, \delta_2), \dots\}$$

which lives on  $\mathsf{X} \times \{0, 1\}$ . Given  $X_i = x$ , then  $\delta_i$  and  $X_{i+1}$  are found by

- 
1. Simulate  $\delta_i \sim \text{Bernoulli}(s(x))$
  2. If  $\delta_i = 1$ , simulate  $X_{i+1} \sim Q(\cdot)$ ; otherwise  $X_{i+1} \sim R(x, \cdot)$ .
- 

Two things are apparent from this construction. First, by (9) the marginal sequence  $\{X_n\}$  has Markov transition kernel given by  $P$ . Second, the set of  $n$  for which  $\delta_{n-1} = 1$ , called *regeneration times*, represent times at which the chain probabilistically restarts itself in the sense that  $X_n \sim Q(\cdot)$  doesn't depend on  $X_{n-1}$ .

The main practical impediment to the use of regenerative simulation would appear to be the means to simulate from the residual kernel  $R(\cdot, \cdot)$ , defined at (10). Interestingly, as shown by Mykland et al. (1995), this is essentially a non-issue, as there is an equivalent update rule for the split chain which does not depend on  $R$  at all. Given  $X_k = x$ , find  $X_{k+1}$  and  $\delta_k$  by

- 
1. Simulate  $X_{k+1} \sim P(x, \cdot)$
  2. Simulate  $\delta_k \sim \text{Bernoulli}(r(X_k, X_{k+1}))$  where

$$r(x, y) = \frac{s(x)Q(dy)}{P(x, dy)}. \quad (11)$$


---

RS has received considerable attention in the case where either a Gibbs sampler or a full-dimensional Metropolis-Hastings sampler is employed. In particular, Mykland et al. (1995) give recipes for establishing minorization conditions as in (8), which have been implemented in several practically relevant statistical models; see e.g. Hobert et al. (2006); Jones et al. (2006); Jones and Hobert (2001); Roy and Hobert (2007).

Suppose we start  $X'$  with  $X_0 \sim Q$ ; one can always discard the draws preceding the first regeneration to guarantee this, but it is frequently easy to draw directly from  $Q$  (Hobert et al., 2002; Mykland et al., 1995). We will write  $E_Q$  to denote expectation when the split chain is started with  $X_0 \sim Q$ . Let  $0 = \tau_0 < \tau_1 < \tau_2 < \dots$  be the regeneration times so

that  $\tau_{t+1} = \min \{k > \tau_t : \delta_{i-1} = 1\}$ . Assume  $X'$  is run for  $R$  tours so that the simulation is terminated the  $R$ th time that a  $\delta_i = 1$ . Let  $\tau_R$  be the total length of the simulation and  $N_t = \tau_t - \tau_{t-1}$  be the length of the  $t$ th tour. Recall that  $g : \mathbf{X} \rightarrow \mathbb{R}$ ,  $Y_i = g(X_i)$ , and  $U_i(y) = I(Y_i \leq y)$ . Also, define

$$S_t = S_t(y) = \sum_{i=\tau_{t-1}}^{\tau_t-1} U_i(y) \quad t = 1, \dots, R.$$

The split chain construction ensures that the pairs  $(N_t, S_t)$  are independent and identically distributed. For each  $y \in \mathbb{R}$  set

$$\Gamma(y) = E_Q \left[ (S_1 - F_V(y)N_1)^2 \right] / [E_Q(N_1)]^2.$$

Let  $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  satisfy  $\lim_{R \rightarrow \infty} h(\tau_R) / \sqrt{\tau_R} = 0$ . Set  $j = \tau_R q + h(\tau_R)$  and consider estimating  $\xi_q$  with  $Y_{\tau_R(j)}$ , that is, the  $j$ th order statistic of  $Y_1, \dots, Y_{\tau_R}$ . The proof of the following CLT is given in Appendix A.

**Theorem 2.** *If  $X$  is polynomially ergodic of order  $m \geq 2$ , then  $E_Q N_r^2 < \infty$  and  $E_Q S_r^2 < \infty$  and hence  $\Gamma(y)$  exists. If, in addition,  $F_V$  has a density  $f_V$  positive and bounded in some neighborhood of  $\xi_q$  with  $f_V$  differentiable at  $\xi_q$ , then as  $R \rightarrow \infty$*

$$\sqrt{R} (Y_{\tau_R(j)} - \xi_q) \xrightarrow{d} N(0, \Gamma(\xi_q) / f_V^2(\xi_q)).$$

Notice  $\hat{\xi}_{\tau_R, q}$  (recall (1)) requires  $j$  such that  $0 < j - \tau_R q \leq 1$  and hence we have the following corollary.

**Corollary 1.** *If  $X$  is polynomially ergodic of order  $m \geq 2$ ,  $F_V$  has a density  $f_V$  positive and bounded in some neighborhood of  $\xi_q$  with  $f_V$  differentiable at  $\xi_q$ , then as  $R \rightarrow \infty$*

$$\sqrt{R} (\hat{\xi}_{\tau_R, q} - \xi_q) \xrightarrow{d} N(0, \Gamma(\xi_q) / f_V^2(\xi_q)).$$

To obtain a Monte Carlo standard error we need to estimate  $\gamma_R^2(\xi_q) := \Gamma(\xi_q) / f_V^2(\xi_q)$ . Just as above we substitute  $\hat{\xi}_{\tau_R, q}$  for  $\xi_q$  and separately consider  $\Gamma(\hat{\xi}_{\tau_R, q})$  and  $f_V(\hat{\xi}_{\tau_R, q})$ . Of course, we can handle estimating  $f_V(\hat{\xi}_{\tau_R, q})$  exactly as before, so all we need to concern ourselves with is estimation of  $\Gamma(\hat{\xi}_{\tau_R, q})$ .

We can recognize  $\Gamma(y)$  as the variance of an asymptotic Normal distribution. Let  $\bar{N}$  denote the average tour length, that is,  $\bar{N} := R^{-1} \sum_{t=1}^R N_t$ , and similarly define  $\bar{S} := R^{-1} \sum_{t=1}^R S_t$ . Now we can estimate  $F_V$  with  $\hat{F}_R := \bar{S} / \bar{N}$  since

$$\hat{F}_R(y) = \frac{\sum_{t=1}^R S_t(y)}{\sum_{t=1}^R N_t} = \frac{1}{\tau_R} \sum_{i=1}^{\tau_R} U_i(y) \quad (12)$$

and hence by the SLLN, as  $R \rightarrow \infty$ ,  $\hat{F}_R(y) \rightarrow F_V(y)$  for each fixed  $y$ . For strictly stationary chains, Tucker (1959) proves this convergence is uniform, i.e. a generalization of the Glivenko-Cantelli theorem; see also Rao (1962) and Tweedie (1977).

From Theorem 2 we have that  $E_Q N_r^2 < \infty$  and  $E_Q S_r^2 < \infty$  and hence we can easily obtain a central limit theorem for  $\hat{F}_R(y)$ . Specifically, for each  $y \in \mathbb{R}$ , as  $R \rightarrow \infty$ ,

$$\sqrt{R} \left( \hat{F}_R(y) - F_V(y) \right) \xrightarrow{d} N(0, \Gamma(y)) .$$

A natural estimator of  $\Gamma(y)$  (Hobert et al., 2002; Jones et al., 2006) is given by

$$\hat{\Gamma}_R(y) = \frac{1}{R\bar{N}^2} \sum_{t=1}^R (S_t - \hat{F}_R(y)N_t)^2 .$$

Letting  $\hat{f}_V(\hat{\xi}_{\tau_R, q})$  denote the density estimator of  $f_V(\hat{\xi}_{\tau_R, q})$  we can estimate  $\gamma_R^2(\xi_q)$  with

$$\hat{\gamma}_R^2(\hat{\xi}_{\tau_R, q}) := \frac{\hat{\Gamma}(\hat{\xi}_{\tau_R, q})}{\hat{f}_V(\hat{\xi}_{\tau_R, q})} .$$

If  $t_{R-1, \alpha/2}$  is an appropriate quantile from a Student's  $t$  distribution with  $R - 1$  degrees of freedom a  $100(1 - \alpha)\%$  confidence interval for  $\xi_q$  is

$$\hat{\xi}_{\tau_R, q} \pm t_{R-1, \alpha/2} \frac{\hat{\gamma}_R(\hat{\xi}_{\tau_R, q})}{\sqrt{R}} . \quad (13)$$

## 2.2 Alternative Methods

It is natural to consider the utility of bootstrap methods for estimating quantiles and the Monte Carlo error. Indeed, there has been a substantial amount of work on using bootstrap methods for stationary time-series, some of which is appropriate for use in MCMC; see e.g., Bertail and Cléménçon (2006), Bühlmann (2002), Carlstein (1986), Datta and McCormick (1993), Politis (2003).

One of the simplest approaches is to use a non-overlapping block resampling scheme, sometimes called 'Carlstein's method' (Carlstein, 1986). The basic idea is to split the Markov chain,  $X$ , into  $a$  non-overlapping blocks of length  $b$ , where we suppose the algorithm is run for a total of  $n = ab$  iterations where  $a = a_n$  and  $b = b_n$  are functions of  $n$ . Then sample the blocks independently with replacement and with equal probability and put these blocks together (end to end) to form a new series. From this new series obtain an estimate of  $\xi_q$ . Repeating this procedure  $p$  times results in  $p$  independent and identically distributed estimates. We can then appeal to classical bootstrap results to estimate the Monte Carlo error of  $\hat{\xi}_{n, q}$ .



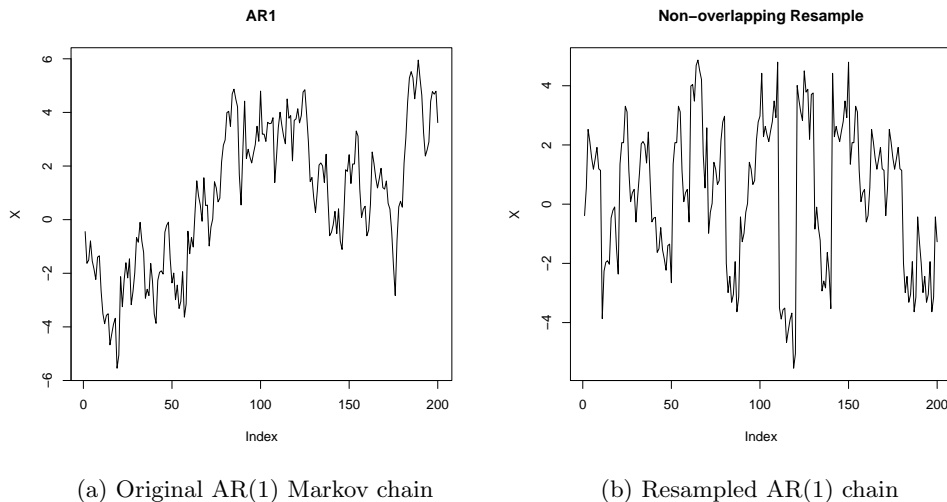


Figure 1: Plots illustrating non-overlapping bootstrap resampling for time-series data using the AR(1) model.

While the simplicity of this method is appealing, it does not seem to work well in MCMC settings. For illustration purposes, consider the first order autoregressive process

$$X_i = \rho X_{i-1} + \epsilon_i \quad \text{for } i = 1, 2, \dots,$$

where  $\epsilon_i$  is an i.i.d.  $N(0, \tau^2)$  for  $i = 1, 2, \dots$ . Figure 1a shows a plot of  $\{X_1, X_2, \dots, X_{200}\}$  with  $\rho = .95$  for one realization of this Markov chain. Using this data, Figure 1b shows a single non-overlapping bootstrap replicate with 10 batches, each of length 20. We can see the method does not retain the dependence structure of the original Markov chain.

In general, resampling results in sequences that are less dependent than the original. As in the AR(1) example, a chain with strong autocorrelation, block bootstrap resampling can result in unrepresentative samples. There are also examples that “can lead to catastrophically bad resampling approximations” (Davison and Hinkley, 1997, p. 397). Further, we found block resampling to be prohibitively computationally intensive.

Another alternative is the subsampling bootstrap (Politis et al., 1999). The basic idea is to split  $X$  into  $n - b + 1$  overlapping blocks of length  $b$ . We then estimate  $\xi_q$  over each block resulting in  $n - b + 1$  estimates. These estimates can be used to estimate the asymptotic variance from Theorem 1. While this approach is better in terms of computational effort than the moving blocks bootstrap it dwarfs the effort required by the methods of the current paper. Moreover, the resulting confidence intervals suffer slightly worse performance in finite samples. See Flegal (2012a) and Flegal and Jones (2011) for more on subsampling bootstrap

methods in the context of MCMC.

### 3 Examples

In this section we investigate the finite-sample performance of the confidence intervals for  $\xi_q$  defined at (7) and (13). While our two examples are quite different, the simulation studies were conducted using a common methodology. In each case we perform many independent replications of the MCMC sampler. Each replication was performed for a fixed number of regenerations, then both the BM- and RS-based confidence intervals were constructed on the same MCMC output. For the BM-based interval we always used  $b_n = n^{1/2}$ , which is the default setting in the `mcmcse` R package. In order to estimate coverage probabilities we require the true values of the quantiles of interest. These are available in only one of our examples. In the other example we estimate the truth with an independent long run of the MCMC sampler. The details are described in the following sections.

#### 3.1 Polynomial target distribution

Jarner and Roberts (2007) studied Markov chain Monte Carlo for heavy-tailed target distributions. A target distribution is said to be *polynomial of order  $r$*  if its density satisfies  $f_V(x) = (l(|x|)/|x|)^{1+r}$ , where  $r > 0$  and  $l$  is a normalized slowly varying function—a particular example is Student’s  $t$ -distribution. We consider estimating quantiles of Student’s  $t$ -distribution  $t(v)$  for degrees of freedom  $v = 3, 5$ , and  $7$ ; the  $t(v)$  distribution is polynomial of order  $v$ . We use a Metropolis random walk algorithm with jump proposals drawn from a  $N(0, \sigma^2)$  distribution. By Proposition 3 of Jarner and Roberts (2007), a Metropolis random walk for a  $t(v)$  target distribution using any proposal kernel with finite variance is polynomially ergodic of order  $v/2$ . Thus the conditions of Theorem 1 are met if  $v \geq 5 + \epsilon$ , while the conditions of Corollary 1 are satisfied for  $v \geq 4$ ; see the first row of Table 1.

We tuned the scale parameter  $\sigma^2$  in the proposal distribution in order to minimize autocorrelation in the resulting chain (second row Table 1); the resulting acceptance rates varied from about 25% for  $t(3)$  with  $\sigma = 6$ , the heaviest tailed target distribution, to about 40% for  $t(7)$  with  $\sigma = 3$ . Regeneration times were identified using the retrospective method of Mykland et al. (1995); see Appendix B for implementation details, and the bottom rows Table 1 for regeneration performance statistics (mean and SD of tour lengths). For each of the  $10^4$  replications and using each of (7) and (13) we computed a 95% confidence interval for  $\xi_q$  for  $q = 0.50, 0.75, 0.90$ , and  $0.95$ . We then compared the observed coverage rate (percentage of the  $10^4$  intervals that indeed contain the true quantile  $\xi_q$ ) with the nominal rate of 95%.

	Target distribution		
	$t(7)$	$t(5)$	$t(3)$
MCSE estimation	BM	RS	
Tuning parameter $\sigma$	3.0	4.0	6.0
Mean tour length	3.91	4.55	5.94
SD of tour lengths	3.50	4.15	7.07

Table 1: *Metropolis random walk on  $t(v)$  target distribution with  $N(0, \sigma^2)$  jump proposals, example of section 3.1. In first row of table “BM” indicates polynomial ergodicity of order  $m \geq 2.5 + \epsilon$ , guaranteeing the conditions of both Theorem 1 and Corollary 1; “RS” indicates  $m \geq 2$ , guaranteeing the conditions of Corollary 1.*

The results are shown in Table 2. There are at least three things to note about the results. First, the empirical coverage rates of BM and RS are comparable, but more closely match the nominal rate for RS than BM. Second, as might be expected, agreement with the nominal coverage rate is closer for estimation of the median than for the tail quantiles  $\xi_{.90}$  and  $\xi_{.95}$ . Third, despite the fact that the assumptions of our theorems do not hold for the  $t(3)$  case, the empirical coverage rates are comparable to, even if slightly worse than, the  $t(7)$  case where the assumptions for both BM and RS are met. Of all three observations it should be noted that the situation is improved by greater simulation effort, ie, running the chain for a greater number of regenerations.

Table 3 shows the mean and standard deviation of the width of the  $10^4$  intervals for the three cases where both BM and RS achieved empirical coverage rates of at least 0.94. We see that intervals are slightly wider on average for RS, while also being less variable than for BM.

Estimating  $\xi_q$  of  $t(v)$  distribution based on Normal Metropolis RW

$R = 500$	$q = .50$		$q = .75$		$q = .90$		$q = .95$	
	BM	RS	BM	RS	BM	RS	BM	RS
$t(7)$	0.942	0.953	0.930	0.943	0.918	0.927	0.905	0.915
$t(5)$	0.939	0.950	0.936	0.945	0.917	0.927	0.902	0.909
$t(3)$	0.939	0.948	0.930	0.943	0.915	0.927	0.902	0.911
$R = 2000$	$q = .50$		$q = .75$		$q = .90$		$q = .95$	
	BM	RS	BM	RS	BM	RS	BM	RS
$t(7)$	0.949	0.955	0.945	0.950	0.937	0.943	0.931	0.936
$t(5)$	0.944	0.951	0.941	0.948	0.938	0.944	0.935	0.942
$t(3)$	0.944	0.949	0.942	0.946	0.939	0.943	0.930	0.937

Table 2: *Empirical coverage rates for nominal 95% confidence intervals for  $\xi_q$ , the  $q$ -quantile of the  $t(v)$  distribution. Based on  $10^4$  replications of  $R$  regenerations of a Metropolis random walk with jump proposals drawn from a Normal distribution. Monte Carlo standard errors are given by  $\sqrt{\hat{p}(1 - \hat{p})/n}$  and fall between .002 and .004.*

$q = 0.50$ and $R = 500$				
Target dist	MCSE Method			
	BM		RS	
$t(7)$	0.250	0.032	0.257	0.023
$t(5)$	0.258	0.033	0.267	0.025
$t(3)$	0.271	0.034	0.283	0.028

$q = 0.50$ and $R = 2000$				
Target dist	MCSE Method			
	BM		RS	
$t(7)$	0.126	0.011	0.128	0.006
$t(5)$	0.130	0.011	0.133	0.007
$t(3)$	0.137	0.012	0.140	0.008

$q = 0.75$ and $R = 2000$				
Target dist	MCSE Method			
	BM		RS	
$t(7)$	0.140	0.013	0.143	0.009
$t(5)$	0.147	0.014	0.149	0.009
$t(3)$	0.161	0.015	0.164	0.011

Table 3: Mean and standard deviation of interval width for 95% confidence intervals for  $\xi_q$ , in  $10^4$  replications of Normal Metropolis random walk with  $R$  regenerations.

### 3.2 Probit Regression

van Dyk and Meng (2001) report data which is concerned with the occurrence of latent membranous lupus nephritis. Let  $y_i$  be an indicator of the disease (1 for present),  $x_{i1}$  be the difference between IgG3 and IgG4 (immunoglobulin G), and  $x_{i2}$  be IgA (immunoglobulin A) where  $i = 1, \dots, 55$ . Suppose

$$\Pr(Y_i = 1) = \Phi(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})$$

and take the prior on  $\beta := (\beta_0, \beta_1, \beta_2)$  to be Lebesgue measure on  $\mathbb{R}^3$ . Roy and Hobert (2007) show that the posterior  $\pi(\beta|y)$  is proper. Our goal is to report a median and an 80% Bayesian credible region for each of the three marginal distributions. Denote the  $q$ th quantile associated with the marginal for  $\beta_j$  as  $\xi_q^{(j)}$  for  $j = 0, 1, 2$ . Then the vector of parameters to be estimated is

$$\Xi = \left( \xi_{.1}^{(0)}, \xi_{.5}^{(0)}, \xi_{.9}^{(0)}, \xi_{.1}^{(1)}, \xi_{.5}^{(1)}, \xi_{.9}^{(1)}, \xi_{.1}^{(2)}, \xi_{.5}^{(2)}, \xi_{.9}^{(2)} \right).$$

We will sample from the posterior using the PX-DA algorithm of Liu and Wu (1999), which Roy and Hobert (2007) prove is geometrically ergodic. For a full description of this algorithm in the context of this example see Flegal and Jones (2010) or Roy and Hobert (2007).

We now turn our attention to comparing coverage probabilities for estimating elements of  $\Xi$  based on the confidence intervals at (7) and (13) for BM and RS, respectively. We calculated a precise estimate from a long simulation of the PX-DA chain and declared the observed quantiles to be the truth—see Table 4. Roy and Hobert (2007) implement RS for this example and we use their settings exactly with 25 regenerations. This procedure was repeated for 1000 independent replications resulting in a mean simulation effort of 3.89E5 (2400). The resulting coverage probabilities can be found in Table 5. Notice that for the interval at (7) all the coverage probabilities are within two MCSEs of the nominal 0.95 level. However, for RS only 7 of the 9 investigated settings are within two MCSEs of the nominal level. In addition, we can see all the results using RS are below the nominal 0.95 level.

Table 6 gives the empirical mean and standard deviation of the half-width of both the BM-based and RS-based confidence intervals. Notice that the RS-based interval is always wider than the BM-based interval.

$q$	0.1	0.5	0.9
$\beta_0$	-5.348 (7.21E-03)	-2.692 (4.00E-03)	-1.150 (2.32E-03)
$\beta_1$	3.358 (4.79E-03)	6.294 (7.68E-03)	11.323 (1.34E-02)
$\beta_2$	1.649 (2.98E-03)	3.575 (5.02E-03)	6.884 (8.86E-03)

Table 4: Summary for Probit Regression example of calculated “truth”. These calculations are based on 9E6 iterations where the MCSEs are calculated using a BM procedure.

	$q$	0.1	0.5	0.9
$\beta_0$	BM	0.956	0.948	0.945
	RS	0.942	0.936	0.934
$\beta_1$	BM	0.948	0.943	0.948
	RS	0.942	0.936	0.934
$\beta_2$	BM	0.949	0.950	0.950
	RS	0.938	0.940	0.937

Table 5: Summary for estimated coverage probabilities for Probit Regression example. CIs reported have 0.95 nominal level with standard errors equal to  $\sqrt{\hat{p}(1-\hat{p})/1000}$ , resulting in MCSEs between 6.5E-3 and 7.9E-3.

	$q$	0.1	0.5	0.9
$\beta_0$	BM	0.0671 (0.007)	0.0377 (0.004)	0.0222 (0.002)
	RS	0.0676 (0.015)	0.0384 (0.008)	0.0226 (0.005)
$\beta_1$	BM	0.0453 (0.005)	0.0720 (0.007)	0.1260 (0.013)
	RS	0.0459 (0.010)	0.0733 (0.016)	0.1270 (0.028)
$\beta_2$	BM	0.0287 (0.003)	0.0474 (0.005)	0.0825 (0.009)
	RS	0.0292 (0.006)	0.0481 (0.010)	0.0831 (0.018)

Table 6: Summary of observed CI half-widths for Probit Regression example with observed standard deviation.

## 4 Discussion

We have established quantile central limit theorems appropriate for use in MCMC settings and considered two methods for estimating the variance of the asymptotic Normal distributions resulting in the interval estimators given at (7) and (13). We also studied the finite-sample properties of these intervals in the context of two examples.

While the regularity conditions for (7) are slightly stronger than those for (13), the intervals were comparable in terms of their finite sample properties. Although minorization conditions are often nearly trivial to establish, they are often seen as a substantial barrier to the implementation of RS. Indeed, it is straightforward to implement the BM-based approach in general software—see the `mcmcse` R package (Flegal, 2012b)—while RS requires a problem-specific approach. For this reason we would recommend the interval at (7).

## A Proof of Theorem 2

We must first show that  $E_Q N_t^2 < \infty$  and  $E_Q S_t^2 < \infty$ . Note that  $0 \leq S_t \leq N_t$  for all  $y \in \mathbb{R}$ , and hence  $E_Q S_t^2 \leq E_Q N_t^2 < \infty$ . Then the first part of Theorem 2 obtains by the following result.

**Lemma 1.** (*Jones and Latuszyński, 2012*) *If  $X$  is polynomially ergodic of order  $m \geq 2$ , then  $E_Q N_r^2 < \infty$ .*

We will also require a preliminary result before proceeding with the rest of the proof; see Iglehart (1976) for related material.

**Lemma 2.** *If  $X$  is polynomially ergodic of order  $m \geq 2$  and  $F_V$  has a density  $f_V$  positive and bounded in some neighborhood of  $\xi_q$ , then  $\Gamma(y)$  is continuous at  $\xi_q$ .*

*Proof.* Denote the limit from the right and left as  $\lim_{y \rightarrow x^+}$  and  $\lim_{y \rightarrow x^-}$ , respectively. From the assumption on  $F_V$  it is clear that

$$\lim_{y \rightarrow \xi_q^+} F_V(y) = \lim_{y \rightarrow \xi_q^-} F_V(y) . \quad (14)$$

Let  $Z_1(y) = S_1(y) - F_V(y)N_1$  and note  $E_Q [Z_1(y)] = 0$  since Hobert et al. (2002) show

$$E_Q S_1(y) = F_V(y)E_Q N_1 \text{ for all } y \in \mathbb{R} . \quad (15)$$



Equations (14) and (15) yield  $E_Q \left[ \lim_{y \rightarrow \xi_q^+} S_1(y) \right] = E_Q \left[ \lim_{y \rightarrow \xi_q^-} S_1(y) \right]$ . The composition limit law and (14) result in

$$E_Q \left[ \lim_{y \rightarrow \xi_q^+} Z_1(y)^2 \right] = E_Q \left[ \lim_{y \rightarrow \xi_q^-} Z_1(y)^2 \right]. \quad (16)$$

What remains to show is that the limit of the expectation is the expectation of the limit. Notice that  $0 < S_1(y) \leq N_1$  for all  $y \in \mathbb{R}$  and

$$|Z_1(y)| = |S_1(y) - F_V(y)N_1| \leq S_1(y) + N_1 \leq 2N_1,$$

which implies  $E_Q [Z_1(y)^2] \leq 4E_Q N_1^2$ . Since  $X$  is polynomially ergodic of order  $m \geq 2$  we have that  $E_Q N_1^2 < \infty$  by Lemma 1 and, the dominated convergence theorem gives, for any finite  $x$ ,

$$\lim_{y \rightarrow x} E_Q [Z_1(y)^2] = E_Q \left[ \lim_{y \rightarrow x} Z_1(y)^2 \right].$$

Finally, from the above fact and (16) we have  $\lim_{y \rightarrow \xi_q^+} E_Q [Z_1(y)^2] = \lim_{y \rightarrow \xi_q^-} E_Q [Z_1(y)^2]$ , and hence  $E_Q [Z_1(y)^2]$  is continuous at  $\xi_q$  implying the desired result.  $\square$

We now return to the proof of Theorem 2. Notice

$$\begin{aligned} P \left( \sqrt{R} (Y_{\tau_R(j)} - \xi_q) \leq y \right) &= P \left( Y_{\tau_R(j)} \leq \xi_q + y/\sqrt{R} \right) \\ &= P \left( \sum_{k=1}^{\tau_R} I\{Y_k \leq \xi_q + y/\sqrt{R}\} \geq j \right) \\ &= P \left( \sum_{k=1}^{\tau_R} \left[ I\{Y_k \leq \xi_q + y/\sqrt{R}\} - F_V \left( \xi_q + y/\sqrt{R} \right) \right] \right. \\ &\quad \left. \geq j - \tau_R F_V \left( \xi_q + y/\sqrt{R} \right) \right) \\ &= P \left( \frac{\sqrt{R}}{\tau_R} \sum_{k=1}^{\tau_R} W_{R,k} \geq s_R \right), \end{aligned}$$

where

$$W_{R,k} = I\{Y_k \leq \xi_q + y/\sqrt{R}\} - F_V \left( \xi_q + y/\sqrt{R} \right), \quad i = 1, \dots, \tau_R,$$

and

$$s_R = \frac{\sqrt{R}}{\tau_R} \left( j - \tau_R F_V \left( \xi_q + y/\sqrt{R} \right) \right).$$

First, consider the  $s_R$  sequence. A Taylor series expansion yields

$$F_V \left( \xi_q + y/\sqrt{R} \right) = F_V (\xi_q) + \frac{y}{\sqrt{R}} f_V' (\xi_q) + \frac{y^2}{2R} f_V'' (\zeta) \quad (17)$$

where  $\zeta$  is between  $\xi_q$  and  $\xi_q + y/\sqrt{R}$ . The definition of  $j$  and (17) result in

$$\begin{aligned}
s_R &= \frac{\sqrt{R}}{\tau_R} \left( j - \tau_R F_V(\xi_q) - \frac{y\tau_R}{\sqrt{R}} f_V(\xi_q) - \frac{y^2\tau_R}{2R} f'_V(\zeta) \right) \\
&= \frac{\sqrt{R}}{\tau_R} \left( \tau_R q + h(\tau_R) - \tau_R q - \frac{y\tau_R}{\sqrt{R}} f_V(\xi_q) - \frac{y^2\tau_R}{2R} f'_V(\zeta) \right) \\
&= -y f_V(\xi_q) + \frac{h(\tau_R)\sqrt{R}}{\tau_R} - \frac{y^2}{2\sqrt{R}} f'_V(\zeta) \\
&= -y f_V(\xi_q) + \frac{h(\tau_R)}{\sqrt{N}\sqrt{\tau_R}} - \frac{y^2}{2\sqrt{R}} f'_V(\zeta)
\end{aligned} \tag{18}$$

and hence  $\lim_{R \rightarrow \infty} s_R = -y f_V(\xi_q)$  by assumptions on  $F_V$ , properties of  $h(\tau_R)$  and the fact that  $\bar{N} \rightarrow E(N_r)$  a.s. where  $1 \leq E(N_r) < \infty$ .

Second, consider  $W_{R,k}$

$$\frac{\sqrt{R}}{\tau_R [\Gamma(\xi_q + y/\sqrt{R})]^{1/2}} \sum_{k=1}^{\tau_R} W_{R,k} \xrightarrow{d} N(0, 1) .$$

Lemma 2 and the continuous mapping theorem imply

$$\frac{\sqrt{R}}{\tau_R [\Gamma(\xi_q)]^{1/2}} \sum_{k=1}^{\tau_R} W_{R,k} \xrightarrow{d} N(0, 1) . \tag{19}$$

Using (18), (19), and Slutsky's Theorem, we can conclude as  $n \rightarrow \infty$

$$\begin{aligned}
P\left(\sqrt{R}(Y_{\tau_R(j)} - \xi_q) \leq y\right) &= P\left(\frac{\sqrt{R}}{\tau_R [\Gamma(\xi_q)]^{1/2}} \sum_{k=1}^{\tau_R} W_{R,k} \geq \frac{s_R}{[\Gamma(\xi_q)]^{1/2}}\right) \\
&\xrightarrow{d} 1 - \Phi\left\{\frac{-y f_V(\xi_q)}{[\Gamma(\xi_q)]^{1/2}}\right\} = \Phi\left\{\frac{y f_V(\xi_q)}{[\Gamma(\xi_q)]^{1/2}}\right\} ,
\end{aligned}$$

resulting in

$$\sqrt{R}(Y_{\tau_R(j)} - \xi_q) \xrightarrow{d} N\left(0, \frac{\Gamma(\xi_q)}{f_V^2(\xi_q)}\right) .$$

## B Regenerative simulation in example of section 3.1

The minorization condition necessary for RS is, at least in principle, quite straightforward for a Metropolis-Hastings algorithm. Let  $q(x, y)$  denote the proposal kernel density, and  $\alpha(x, y)$  the acceptance probability. Then  $P(x, dy) \geq q(x, y)\alpha(x, y)dy$ , since the right hand side only accounts for *accepted* jump proposals, and the minorization condition is established by finding

$s'$  and  $\nu'$  such that  $q(x, y)\alpha(x, y) \geq s'(x)\nu'(y)$ . By Theorem 2 of Mykland et al. (1995), the probability of regeneration on an *accepted* jump from  $x$  to  $y$  is then given by

$$r_A(x, y) = \frac{s'(x)\nu'(y)}{q(x, y)\alpha(x, y)}.$$

Letting  $\pi$  denote the (possibly unnormalized) target density, we have for a Metropolis random walk

$$\alpha(x, y) = \min \left\{ \frac{\pi(y)}{\pi(x)}, 1 \right\} \geq \min \left\{ \frac{c}{\pi(x)}, 1 \right\} \min \left\{ \frac{\pi(y)}{c}, 1 \right\}$$

for any positive constant  $c$ . Further, for any point  $\tilde{x}$  and any set  $D$  we have

$$q(x, y) \geq \inf_{y \in D} \left\{ \frac{q(x, y)}{q(\tilde{x}, y)} \right\} q(\tilde{x}, y) I_D(y).$$

Together, these inequalities suggest one possible choice of  $s'$  and  $\nu'$ , which results in

$$r_A(x, y) = I_D(y) \times \frac{\inf_{y \in D} \{q(x, y)/q(\tilde{x}, y)\}}{q(x, y)/q(\tilde{x}, y)} \times \frac{\min \{c/\pi(x), 1\} \min \{\pi(y)/c, 1\}}{\min \{\pi(y)/\pi(x), 1\}}. \quad (20)$$

For a  $t(v)$  target distribution,  $\alpha(x, y)$  reduces to

$$\min \left\{ \left( \frac{v + x^2}{v + y^2} \right)^{\frac{v+1}{2}}, 1 \right\} \geq \min \left\{ \left( \frac{v + x^2}{c} \right)^{\frac{v+1}{2}}, 1 \right\} \times \min \left\{ \left( \frac{c}{v + y^2} \right)^{\frac{v+1}{2}}, 1 \right\}$$

and the last component of (20) is given, up to the constant  $c$ , by

$$\left[ \frac{\min \{v + x^2, c\}}{\min \{v + x^2, v + y^2\}} \times \frac{v + y^2}{\max \{v + y^2, c\}} \right]^{\frac{v+1}{2}}.$$

Since this piece of the acceptance probability takes the value 1 whenever  $v + x^2 < c < v + y^2$  or  $v + y^2 < c < v + x^2$ , it makes sense to take  $c$  equal to the median value of  $v + X^2$  under the target distribution.

The choice of  $\tilde{x}$  and  $D$ , and the functional form of the middle component of (20), will of course depend on the proposal distribution. For the Metropolis random walk with Normally distributed jump proposals,  $q(x, y) \propto \exp \left\{ -\frac{1}{2\sigma^2}(y - x)^2 \right\}$ , taking  $D = [\tilde{x} - d, \tilde{x} + d]$  for  $d > 0$  gives

$$\frac{\inf_{y \in D} \{q(x, y)/q(\tilde{x}, y)\}}{q(x, y)/q(\tilde{x}, y)} = \exp \left\{ -\frac{1}{\sigma^2} \{(x - \tilde{x})(y - \tilde{x}) + d|x - \tilde{x}|\} \right\}.$$

For the  $t(v)$  distributions we can take  $\tilde{x} = 0$  in all cases, but the choice of  $d$  should depend on  $v$ . With the goal of maximizing regeneration frequency, we arrived at, by trial and error,  $d = 2\sqrt{v/(v - 2)}$ , or two standard deviations in the target distribution.

## References

- Bertail, P. and Cl  men  on, S. (2006). Regenerative block-bootstrap for Markov chains. *Bernoulli*, 12:689–712.
- B  hlmann, P. (2002). Bootstraps for time series. *Statistical Science*, 17:52–72.
- Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 14:1171–1179.
- Datta, S. and McCormick, W. P. (1993). Regeneration-based bootstrap for Markov chains. *The Canadian Journal of Statistics*, 21:181–193.
- Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.
- Douc, R., Fort, G., Moulines, E., and Soulier, P. (2004). Practical drift conditions for subgeometric rates of convergence. *The Annals of Applied Probability*, 14:1353–1377.
- Flegal, J. M. (2012a). Applicability of subsampling bootstrap methods in Markov chain Monte Carlo. In Wozniakowski, H. and Plaskota, L., editors, *Monte Carlo and Quasi-Monte Carlo Methods 2010* (to appear). Springer-Verlag.
- Flegal, J. M. (2012b). mcmcse: Monte Carlo standard errors for MCMC R package version 0.1. <http://cran.r-project.org/web/packages/mcmcse/index.html>.
- Flegal, J. M., Haran, M., and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23:250–260.
- Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38:1034–1070.
- Flegal, J. M. and Jones, G. L. (2011). Implementing Markov chain Monte Carlo: Estimating with confidence. In Brooks, S., Gelman, A., Jones, G., and Meng, X., editors, *Handbook of Markov Chain Monte Carlo*, pages 175–197. Chapman & Hall/CRC Press.
- Fort, G. and Moulines, E. (2000). V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statistics and Probability Letters*, 49:401–410.
- Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Processes and their Applications*, 103:57–99.

- Geyer, C. J. (1996). Estimation and optimization of functions. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. E., editors, *Markov chain Monte Carlo in practice*, pages 241–258. Chapman & Hall, Boca Raton.
- Geyer, C. J. (2011). Introduction to Markov chain Monte Carlo. In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Press, London.
- Hobert, J. P. (2011). The data augmentation algorithm: Theory and methodology. In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Press, London.
- Hobert, J. P., Jones, G. L., Presnell, B., and Rosenthal, J. S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, 89:731–743.
- Hobert, J. P., Jones, G. L., and Robert, C. P. (2006). Using a Markov chain to construct a tractable approximation of an intractable probability distribution. *Scandinavian Journal of Statistics*, 33(1):37–51.
- Iglehart, D. L. (1976). Simulating stable stochastic systems, VI: Quantile estimation. *Journal of the ACM*, 23:347–360.
- Jarner, S. F. and Roberts, G. O. (2002). Polynomial convergence rates of Markov chains. *Annals of Applied Probability*, 12:224–247.
- Jarner, S. F. and Roberts, G. O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scandinavian Journal of Statistics*, 24:101–121.
- Jarner, S. F. and Tweedie, R. L. (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli*, 9:559–578.
- Johnson, A. A., Jones, G. L., and Neath, R. C. (2011). Component-wise Markov chain Monte Carlo. *Preprint*.
- Jones, G. L. (2004). On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320.
- Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101:1537–1547.
- Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, 16:312–334.

- Jones, G. L. and Łatuszyński, K. (2012). Output analysis in Markov chain Monte Carlo. Technical report, University of Minnesota, School of Statistics.
- Jones, G. L., Roberts, G. O., and Rosenthal, J. S. (2012). Convergence of conditional Metropolis-Hastings samplers, with an application to inference for discretely-observed diffusions. Technical report, University of Minnesota, School of Statistics.
- Kim, T. Y. and Lee, S. (2005). Kernel density estimator for strong mixing processes. *Journal of Statistical Planning and Inference*, 133(2):273–284.
- Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer, New York.
- Liu, J. S. and Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94:1264–1274.
- Mengersen, K. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24:101–121.
- Mykland, P., Tierney, L., and Yu, B. (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, 90:233–241.
- Politis, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*, 18:219–230.
- Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer-Verlag Inc.
- Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 33:659–680.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York, second edition.
- Roberts, G. O. and Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83:95–110.
- Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4:185–207.
- Roy, V. and Hobert, J. P. (2007). Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *Journal of the Royal Statistical Society, Series B*, 69(4):607–623.
- Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22:1701–1762.

- Tucker, H. G. (1959). A generalization of the Glivenko-Cantelli theorem. *The Annals of Mathematical Statistics*, 30:828–830.
- Tweedie, R. L. (1977). Modes of convergence of Markov chain transition probabilities (MR V56 16794). *Journal of Mathematical Analysis and Applications*, 60:280–291.
- van Dyk, D. A. and Meng, X.-L. (2001). The art of data augmentation (with discussion). *Journal of Computational and Graphical Statistics*, 10:1–50.
- Yoshihara, K.-i. (1995). The Bahadur representation of sample quantiles for sequences of strongly mixing random variables. *Statistics & Probability Letters*, 24:299–304.
- Yu, B. (1993). Density estimation in the  $L^\infty$  norm for dependent data with applications to the Gibbs sampler. *The Annals of Statistics*, 21:711–735.